INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 21, Issue: 4, Month: August, Year: 2026
Article Number: 7599, https://doi.org/10.15837/ijccc.2026.4.7599

Explainable Early Warning Systems for Student Dropout Prediction: Insights from a Bibliometric Analysis

D. A. Radulescu*, V. E. Balas

Dan Andrei Radulescu*
Doctoral School of Systems Engineering
Petroleum-Gas University of Ploiesti
Bd. Bucuresti 39, 100680 Ploiesti, Romania
*Corresponding author: dan.radulescu@upg-ploiesti.ro

Valentina Emilia Balas
1. Faculty of Engineering
Aurel Vlaicu University of Arad, 310032 Arad, Romania
2. Academy of Romanian Scientists, Bucharest, Romania
balas@drbalas.ro

Abstract

Student dropout and weak academic performance are among the most costly problems higher education institutions face, and their early prediction has become an active research field. This paper maps that field through a bibliometric analysis of 352 Scopus-indexed documents published in English between 2014 and 2025, selected from 446 records through a documented screening. Author-keyword co-occurrence analysis in VOSviewer, after thesaurus-based cleaning, yielded 39 recurrent keywords organized into four clusters: an applied core around machine learning and dropout prediction in higher education; classification algorithms and deep-learning architectures; the older educational-data-mining vocabulary; and academic performance, retention and class imbalance. A temporal overlay shows the field moving from generic classification and data mining toward deep learning, gradient boosting and explainable artificial intelligence (XAI). Against the four gaps this map exposes — interpretability, transferability, fairness and deployment — we propose X-EWS, a five-layer explainable early-warning framework spanning data acquisition to intervention and governance, in which each gap is answered by a specific architectural choice. A proof-of-concept instantiation on a public benchmark of 4,424 students illustrates that the framework's predictive, explanatory and audit layers can be realized with standard, auditable components.

Keywords: student dropout prediction, machine learning, bibliometric analysis, explainable artificial intelligence, early warning system.

1 Introduction

Universities today generate and store large amounts of administrative and learning data almost as a by-product of their daily operation. Student information systems, learning management platforms, examination and library services leave a continuous trail of records, and the expectation that this trail should inform decisions about retention, timely graduation and the overall quality of teaching has become difficult to ignore. Within the European Higher Education Area, quality assurance is increasingly framed as evidence-based governance rather than periodic reporting, and predicting which students are likely to drop out or to underperform, early enough for an intervention to matter, is central to that agenda.

The problem itself is old, but the way it is approached has changed considerably. Early efforts relied on descriptive reporting and on classical statistics applied to small, single-program samples. Over the last decade the field has shifted toward learning analytics and educational data mining, where machine learning models are trained to assign individual risk scores from heterogeneous academic and behavioral signals. The volume of work produced in this direction has grown quickly, and several systematic reviews have already tried to organize it [2, 6, 25]. What is less common is a quantitative, structural view of the literature itself: which themes hold the field together, which methods dominate it, and the way its vocabulary has moved over time.

Bibliometric analysis offers exactly this kind of view. Rather than summarizing individual studies, it treats the published corpus as a network and lets recurring terms and their co-occurrences expose the underlying thematic structure. Drawing on the bibliographic data available in Scopus, we performed a topic search built around the keywords *academic performance*, *dropout* and *prediction*, and analyzed the resulting corpus in VOSviewer. The study is guided by four research questions:

- **RQ1.** What are the principal thematic clusters, or research fronts, in the literature on machine learning for student dropout and academic performance prediction in higher education?
- **RQ2.** How has the field's vocabulary and methodological focus evolved over the 2014-2025 period?
- **RQ3.** Which gaps does the bibliometric structure reveal in how these models are explained, transferred across institutions, audited for fairness, and put into actual use?
- **RQ4.** How can these findings inform the design of an explainable early-warning system for higher education?

The first three questions are answered through the bibliometric analysis; the fourth motivates the conceptual framework proposed in Section 8. The paper therefore does two things. It gives a structural and temporal map of the field, and from the gaps that map exposes it builds a five-layer architecture for an explainable early-warning system.

The remainder of the paper is organized as follows. Section 2 reviews representative studies. Section 3 describes the data source and the methodological pipeline. Section 4 gives the clustering results and Section 5 takes up each cluster in turn, while Section 6 reads the same map temporally. The findings and the gaps they expose are discussed in Section 7. Section 8 sets out the explainable early-warning system (X-EWS) framework, Section 9 illustrates it in a proof-of-concept instantiation on a public benchmark, Section 10 draws together the theoretical and practical contributions, and Section 11 concludes.

2 Literature review

The prediction of academic outcomes through data-driven methods has matured into a recognizable research tradition, and a useful entry point is the body of systematic reviews that periodically take stock of it. Albreiki, Zaki and Alashwal surveyed the educational data mining literature on at-risk and dropout students between 2009 and 2021, noting that most studies draw either on institutional databases or on online-platform logs, and that machine learning consistently improves the identification of students in difficulty [2]. Rastrollo-Guerrero and colleagues reached a similar conclusion from a methodological angle, reviewing almost seventy studies and grouping the techniques into machine learning, collaborative filtering, recommender systems and neural networks [25]. Batool and co-authors went further in scale, comparing roughly 260 studies over two decades; they report that artificial neural networks and random forests are the most frequently used algorithms and that students' academic records and demographic factors are the best attributes for predicting performance [6]. De-La-Cruz et al. and Alalawi et al. confirm the same overall picture, with supervised learning, and in particular support vector machines, random forests and neural networks, clearly dominant [1, 9].

A recurring concern in the empirical work is the imbalanced nature of the data. In any cohort far fewer students drop out than persist, and a naïve classifier tends to optimize overall accuracy at the expense of the minority class that actually matters. Thammasiri and colleagues addressed this directly, showing on a large institutional dataset that combining support vector machines with the synthetic minority over-sampling technique (SMOTE) gave the best balance, reaching about 90% overall accuracy (90.24% on a ten-fold holdout) while substantially improving recognition of the minority class [36]. Villar and de Andrade revisited the same issue more recently, pairing SMOTE with boosting algorithms and hyperparameter tuning, and found that the gradient-boosting implementations LightGBM and CatBoost outperformed traditional classifiers [39]. Class imbalance thus runs as a methodological thread from the earliest to the most recent contributions.

Several studies are best read as demonstrations that machine learning can be put to work in real institutional settings. Burgos et al. analyzed historical course-grade data with logistic regression and, importantly, translated the resulting model into a tutoring action plan; relative to previous years without any prevention mechanism, the dropout rate in the affected courses fell by 14% [7]. Rovira, Puertas and Igual built a data-driven system that predicts both grades and dropout intention and accompanies its output with visualizations meant to help tutors act on it [29]. Tan and Shao worked at a very different scale, training neural networks, decision trees and Bayesian networks on more than sixty thousand e-learning students, with decision trees performing best [35]. Lopez Guarin and colleagues modeled the loss of academic status at a Colombian university and showed that adding academic data to admission records markedly improved prediction [19], while Devasia et al. obtained comparable gains from a naïve Bayesian model built on a few hundred students of a campus-based program [11].

The question of *when* to predict has received specific attention. Ortiz-Lozano et al. compared models built at three moments of the first semester and argued that, although early identification is desirable, academic monitoring through the semester improves accuracy enough to challenge a strict "as soon as possible" stance [24]. Alhazmi and Sheneamer combined clustering and classification to flag at-risk students at an early stage from admission scores and first-level course grades [3]. Kabathova and Drlik examined how a deliberately limited and scarce set of course-level features behaves across several classifiers; prediction accuracy ranged between 77% and 93% on the following year's unseen data, and the authors caution that performance metrics must be read carefully on small educational datasets [17].

Online and behavioral data form their own stream. Coussement and colleagues benchmarked a logit leaf model against eight alternatives on more than ten thousand students of a subscription-based provider, arguing that it strikes a good balance between predictive performance and comprehensibility [8]. Umer et al. enriched standard features with process-mining indicators derived from event logs in massive open online courses [37], and He and co-authors proposed a joint recurrent neural network with gated recurrent units (RNN-GRU) to exploit the sequential behavior recorded in the widely used Open University Learning Analytics Dataset (OULAD) [16]. The most recent reviews dedicated to online settings, by Shi et al. and by Rizwan and colleagues, confirm both the growing reliance on behavioral engagement features and the rapid adoption of deep and ensemble architectures over the past five years [27, 32].

Deep learning is the clearest recent inflection. Nabil, Seyam and Abou-Elfetouh used a deep neural network on first-year course grades to identify students at risk of failure with an accuracy of 89%, ahead of several classical baselines on the same data [22]. Vives et al. applied long short-term memory (LSTM) networks, with balancing based on generative adversarial networks (GAN) and SMOTE, to the notoriously difficult Programming Fundamentals course, reporting 98.3% accuracy by the eighth week of the semester [40]. Alnasyan, Basheri and Alassafi reviewed forty-six deep learning studies in virtual learning environments and found deep neural networks (DNN) and convolutional neural network (CNN)-LSTM models to be the most common architectures, with several studies exceeding

90% accuracy and learning-behavior features standing out as the strongest predictors [4]. Stasolla and co-authors broadened the lens to reinforcement learning, mapping how deep and adaptive methods support personalized performance prediction [34].

As models grow more capable, interpretability and fairness move to the foreground. Baranyi, Nagy and Molontay paired strong predictors, deep networks and gradient-boosted trees, with permutation importance and SHAP (SHapley Additive exPlanations) values so that the predictions could be explained; notably, their best deep-learning model (72.4% accuracy, area under the curve [AUC] 0.771) slightly outperformed the XGBoost (eXtreme Gradient Boosting) benchmark while remaining interpretable [5]. Duro et al., reviewing studies published between 2018 and 2025, make the same point at the level of the whole field: transparent traditional models stay dominant precisely because they are easier to explain, and only a minority of studies address fairness, bias or real-world integration [12]. Yu, Lee and Kizilcec confronted the fairness question head-on, asking whether protected attributes should be included in dropout models; on a population of more than ninety thousand students they found that including such attributes barely changed overall performance while marginally improving fairness [41]. Rodríguez-Ortiz and colleagues extend the discussion toward generative AI, observing that while machine learning is well established in learning analytics, generative models remain experimental and carry their own transparency concerns [28].

A final group of studies keeps the focus on explanatory factors rather than on raw predictive power. Kocsis and Molnár synthesized evidence on a very large student population and identified grade point average, accumulated credits and gender as the most consistent predictors, mediated by motivation, self-regulation and prior education [18]. Sinval et al. modeled the interplay of depression, anxiety, stress, engagement and dropout intention in medical students, showing that psychological distress acts on performance indirectly, through engagement [33]. Delogu and colleagues, working with the entire population of Italian bachelor students, confirmed that random forests and gradient boosting are effective early-warning indicators and that first-year performance, family income and high-school background drive the predictions [10]. Monteverde-Suárez et al. reached convergent conclusions in a medical-education setting, where artificial neural networks and naïve Bayes identified at-risk first-year students from prior knowledge and socio-demographic data [21].

These works show a field that has accumulated considerable empirical depth and methodological variety. What the individual studies cannot show is how their themes relate to one another at the level of the whole literature, which is precisely what the bibliometric analysis below sets out to do.

3 Methodology

The analysis is based on bibliographic data retrieved from Scopus. A topic search was issued on 19 May 2026 over the title, abstract and keywords field, combining the central notions of the field; in Scopus syntax the query was `TITLE-ABS-KEY ( academic AND performance AND dropout AND prediction )`. The search returned 446 documents on that date (the same query returned 457 documents when last verified, reflecting the field's continuing growth).

A word on the choice of terms is in order, since any search expression shapes the corpus it returns. The query deliberately targets the intersection of the two outcomes of interest (academic performance, dropout) with the predictive task, rather than enumerating adjacent vocabulary such as *retention*, *attrition*, *student success*, *at-risk students*, *learning analytics* or *academic failure*. Broader OR-expressions of that kind pull in large neighboring literatures, retention policy or descriptive analytics among them, that do not involve prediction at all, and their volume would dilute precisely the intersection this study is about. The restriction, moreover, applies to the query and not to the vocabulary the corpus is allowed to reveal: the resulting keyword map does contain *student retention* (14 occurrences), *at-risk students* (8) and *learning analytics* (27), which indicates that the corpus reaches these adjacent literatures even though the query does not name them. The wording of the query nonetheless remains a scoping decision, and we return to it as a limitation in Section 11.

The corpus was then screened following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) logic shown in Figure 1. Of the 446 records, 20 fell outside the *Article*, *Conference paper* and *Review* types (book chapters, editorials and conference reviews) and were excluded, which left 426. A further 20 non-English records were dropped. Applying the 2014-2025 window then removed 54 more, mostly in-press 2026 items returned by the same query plus a handful of older documents, for a final corpus of 352 documents used to build the term map. The most recent reviews are still discussed narratively in Section 2, even where their formal publication year falls in 2026.

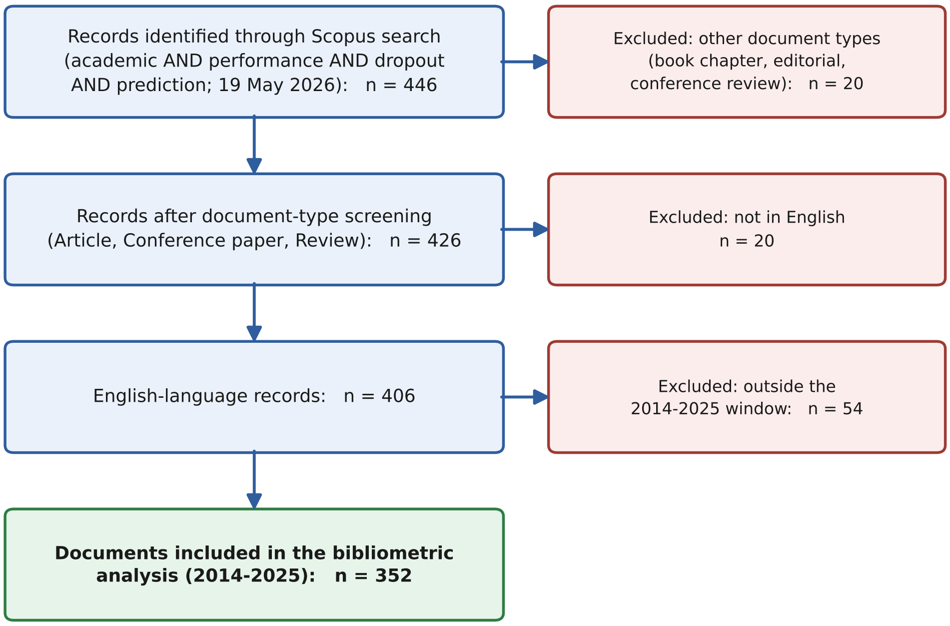

Figure 1: PRISMA flow of the corpus selection (Scopus, 19 May 2026).

Before mapping the term network, we briefly profile the corpus, so that the analysis does not rest on keyword co-occurrence alone. The annual output (Figure 2) climbs steeply across the window, from 5 documents in 2014 to 39 in 2023, 78 in 2024 and 118 in 2025; most of the literature is therefore very recent and the field is still growing.

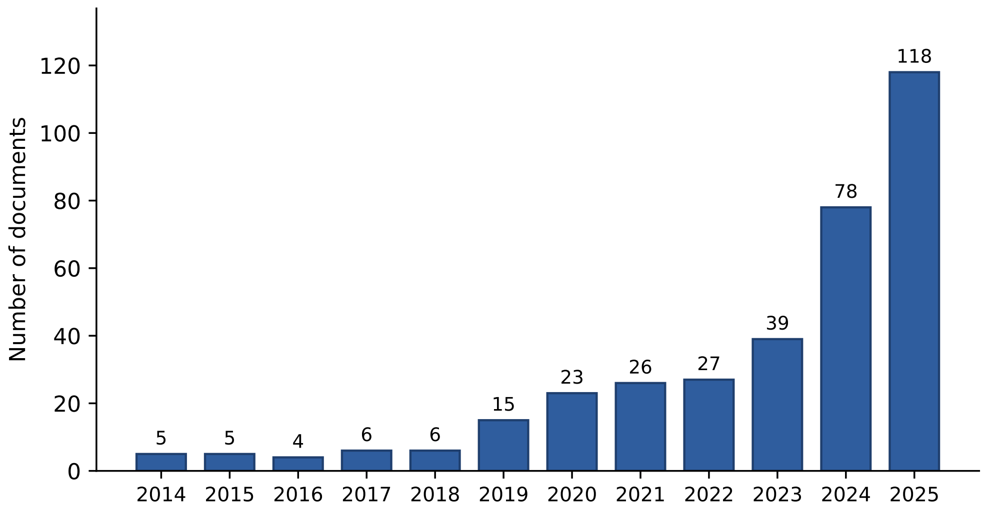

Figure 2: Annual scientific production in the corpus (Scopus, 2014-2025).

The work is scattered across many venues rather than concentrated in a few, which is what one expects of a young interdisciplinary field. Table 1 lists the most productive sources; IEEE Access is well ahead of the next venues. Table 2 lists the most cited documents in the corpus; the ranking is led by review and survey articles [2, 6, 25], as expected in a field mature enough to have produced several syntheses of its own.

Table 1: Most productive sources in the corpus.

Source	Documents
IEEE Access	14
Applied Sciences (Switzerland)	8
Education and Information Technologies	7
Lecture Notes in Computer Science	7
Lecture Notes in Networks and Systems	6
Communications in Computer and Information Science	6
ACM International Conference Proceeding Series	5
Frontiers in Education	3

Table 2: Most cited documents in the corpus.

Study	Year	Source	Cited by
Albreiki et al.	2021	Education Sciences	352
Rastrollo-Guerrero et al.	2020	Applied Sciences	315
Burgos et al.	2018	Computers and Electrical Engineering	198
Nabil et al.	2021	IEEE Access	177
Thammasiri et al.	2014	Expert Systems with Applications	177
Batool et al.	2023	Education and Information Technologies	168
Devasia et al.	2016	Proc. SAPIENCE 2016	148
Tan and Shao	2015	Int. J. Emerging Technologies in Learning	141

The records were then imported into VOSviewer (version 1.6.21), a tool specialized in the construction and visualization of bibliometric networks. A map was built with the unit of analysis set to keyword co-occurrence, using the full counting method and VOSviewer's default association-strength normalization for the link weights. Because author keywords are assigned by the authors themselves and tend to describe the substantive content of a paper more faithfully than automatically indexed terms, the map was constructed on author keywords rather than on index keywords or terms extracted from titles and abstracts. Before counting co-occurrences, a thesaurus file (provided as supplementary material) was applied to clean the term space. Synonymous and morphological variants were merged: for instance, *machine-learning* and *machine learning (ml)* were collapsed into *machine learning*, and *edm*, *random forests* and *xai* were folded into their canonical forms. Generic non-thematic terms, such as the bare word *students*, were removed. After this cleaning the corpus contained 860 distinct author keywords. A minimum co-occurrence threshold of five was set, a common compromise that retains only terms recurrent enough to position reliably on the map while still covering the breadth of the field's vocabulary. Thirty-nine terms met this threshold. On the basis of their co-occurrence strength they were grouped into four clusters, shown in Figure 3 (network visualization) and Figure 4 (density visualization), with their temporal signature in Figure 5 (overlay visualization).

For reproducibility, the thesaurus merged further synonymous and morphological variants beyond the examples above (*deep learning (dl)* and *dl* into *deep learning*; *drop-out* and *drop out* into *dropout*; *support vector machines*, *support vector machine (svm)* and *svm* into *support vector machine*). The complete 39-keyword list, with per-cluster occurrence counts, is given in Section 4. The Scopus export of the 352 records, the complete thesaurus file and the VOSviewer map file are available from the corresponding author as supplementary material. Keyword co-occurrence mapping of this kind has recently been applied to other engineering domains as well, for instance to quality management in the automotive industry in the context of Industry 4.0 [38], which supports its use for structuring the present field.

4 Clustering

The clustering produced four thematic clusters, each corresponding to a different research area within the prediction of dropout and academic performance. Their composition is summarized below; the figures in parentheses give the number of documents in which each keyword appears.

Cluster 1 (13 terms) — machine learning (123), dropout prediction (60), higher education (43), learning analytics (27), student dropout (27), predictive models (21), at-risk students (8), explainable ai (7), gradient boosting (7), predictive analytics (7), feature engineering (5), student success (5), university dropout (5). This is the largest and most connected cluster of the map.

Cluster 2 (10 terms) — deep learning (31), decision tree (19), artificial intelligence (15), neural network (15), education (12), random forest (11), support vector machine (11), logistic regression (8), lstm (7), learning management system (5).

Cluster 3 (10 terms) — educational data mining (91), prediction (45), student performance (35), classification (32), dropout (26), data mining (21), e-learning (7), regression (7), academic success (6), performance prediction (5).

Cluster 4 (6 terms) — academic performance (34), feature selection (20), student retention (14), class imbalance (11), academic achievement (5), academic dropout (5).

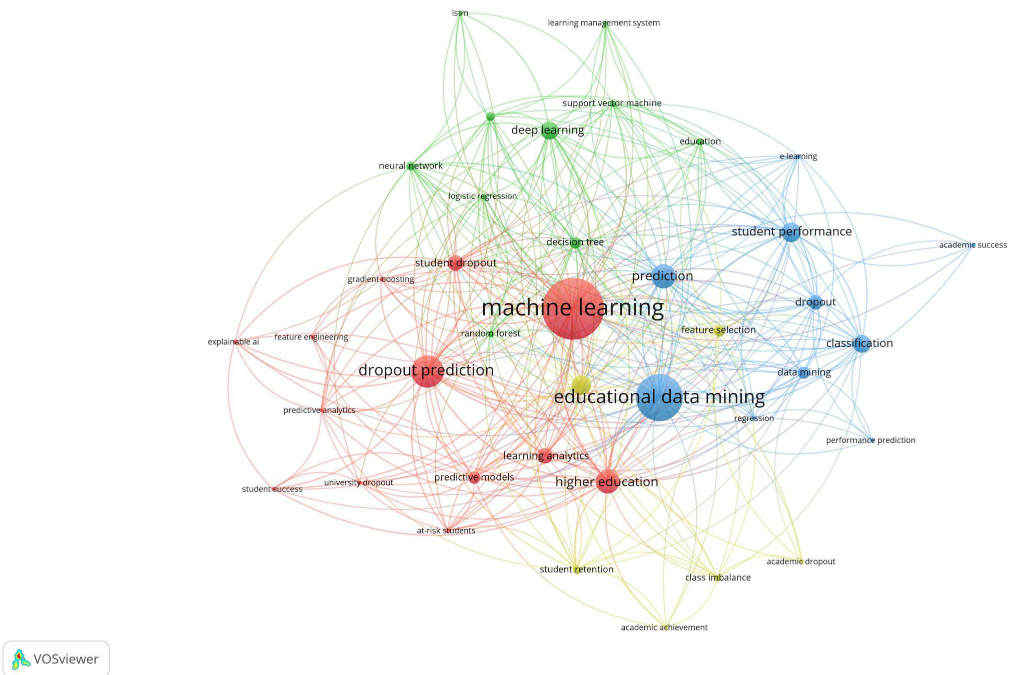

Figure 3: VOSviewer keyword co-occurrence network, colored by cluster.

Read together, the four clusters separate the substance of the problem from the methods used to attack it. Cluster 1 is the applied core, where machine learning meets dropout prediction in higher education. Cluster 4 brings together the outcome variables (academic performance and retention) and the data-quality problems that stand in the way of predicting them. Cluster 3 is the foundational educational-data-mining and prediction vocabulary the field was first built on, while Cluster 2 holds the methods, from classical classifiers through to deep-learning architectures. The qualitative reading that follows develops each cluster in turn.

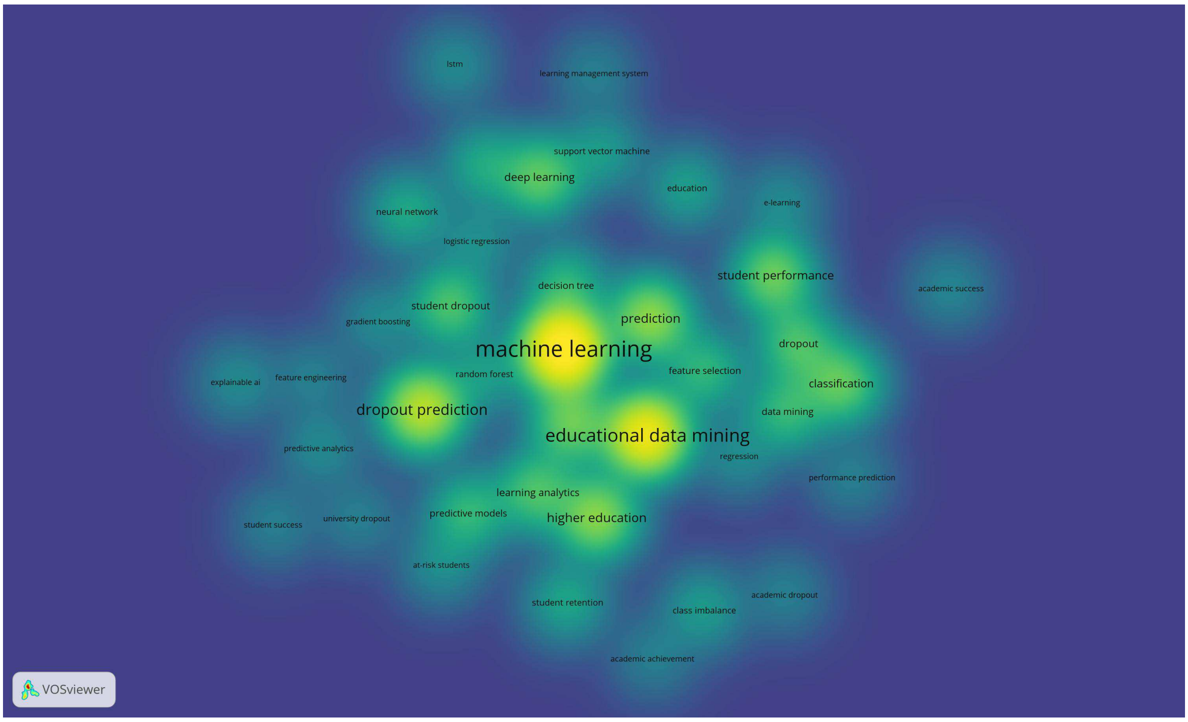

Figure 4: VOSviewer keyword density visualization.

5 Qualitative analysis of the clusters

5.1 Cluster 1: Machine learning for dropout prediction in higher education

The largest and most connected cluster is also the thematic core of the corpus. It binds the central object of study, dropout (increasingly named *dropout prediction*) in *higher education*, to the dominant analytical paradigm, *machine learning*, and to the apparatus of *learning analytics*, *predictive models* and *predictive analytics* that has grown up around it. The co-occurrence of these terms is consistent with the documented convergence of the machine-learning and learning-analytics communities, which now share methods and datasets [2, 28]. Two signals within the cluster are worth singling out. The presence of *at-risk students* as an established term confirms that the field frames its task as the early identification of a vulnerable minority [3, 24]. The appearance of *explainable ai*, *gradient boosting* and *predictive analytics*, all with recent average years, shows that accuracy is no longer the only concern: researchers now want to know how a score was produced before they trust it in practice [5, 12]. The cluster thus sits exactly where the educational problem meets the decision-support ambition that motivates it [7, 8].

5.2 Cluster 2: Classification algorithms and deep-learning architectures

The second cluster contains the methodological apparatus of the field. It places the field's classical classifiers, *decision tree*, *random forest*, *support vector machine* and *logistic regression*, alongside the newer *deep learning*, *neural network* and *lstm*, with *artificial intelligence* as the umbrella term and *education* and *learning management system* marking the data context. That the classical and the deep methods share a single cluster is itself informative: comparative studies routinely benchmark the two families against each other on the same institutional data [6, 17, 22]. Classical models stay in use because they are robust on tabular data and easy to read, whereas deep architectures earn their place mainly when there are sequential or behavioral signals to exploit [4, 34, 40]. The cluster therefore captures not a succession but a coexistence of methods.

5.3 Cluster 3: Educational data mining and the foundational prediction vocabulary

Conceptually prior to the others, the third cluster gathers the vocabulary through which the problem was first framed: *educational data mining*, still the single most frequent author keyword after machine learning, together with *prediction*, *classification*, *data mining*, *dropout*, *regression* and *e-learning*. These are the terms of the period when dropout and performance were treated as generic classification or data-mining tasks, in early e-learning and institutional studies alike [7, 11, 35]. The co-location of *student performance* and *performance prediction* with *classification* and *regression* also captures the dual framing of the dependent variable, sometimes a categorical risk label and sometimes a continuous grade, that still runs through much of the empirical work [19, 25]. As the temporal analysis below shows, this is the oldest cluster, and its terms now act as the shared substrate over which the more specialized vocabulary of the others is built.

5.4 Cluster 4: Academic performance, retention and class imbalance

The smallest cluster is organized around outcomes and the practical obstacles to predicting them. *Academic performance*, *student retention*, *academic achievement* and *academic dropout* are the target variables institutions care about, and they sit beside the two terms that most shape whether a model can predict them well: *feature selection* and *class imbalance*. There is a reason these terms cluster together. Educational datasets combine academic, behavioral and socio-demographic variables of uneven quality, so selecting the informative subset is often what keeps a model usable rather than over-fitted [23, 30]; and since far fewer students leave than stay, dealing with the rarity of the event is a precondition for any accuracy figure that means anything [36, 39]. Studies that predict retention and success directly tend to live here [15, 31].

Synthesis

The clusters together show that dropout and performance prediction has grown into an interdisciplinary area drawing on education sciences, statistics and computer science at once, with a growing and explicit concern for whether its models are useful to institutions and whether they can be interpreted.

6 Temporal analysis

The overlay visualization in Figure 5 colors each term by its average year of appearance across the 2014-2025 window. These are average years, not dates of first occurrence; because publication volume in the field grew steeply after 2020, even terms present since the start of the window carry averages in the 2020-2025 range. Read this way, the map reveals a coherent trajectory that is best described in three phases.

The **foundational layer** is anchored in Cluster 3. Its terms carry the earliest averages: *e-learning* ($\approx$ 2020.3), *classification* ($\approx$ 2020.9), *regression* ($\approx$ 2021.0), *performance prediction* ($\approx$ 2021.4), *dropout* ($\approx$ 2021.7) and *prediction* ($\approx$ 2021.8). They set up the generic framing of the problem as a classification or regression task on educational data, and rather than dropping out of use they simply became background assumptions.

A **consolidation phase** follows, around 2022-2023, as the field matures. *Educational data mining* ($\approx$ 2022.1), *higher education* ($\approx$ 2022.1), *data mining* ($\approx$ 2022.6), *learning analytics* ($\approx$ 2022.5), *student dropout* ($\approx$ 2022.6) and *student retention* ($\approx$ 2022.6) acquire their established meaning, while on the methodological side *decision tree* ($\approx$ 2022.2) and *neural network* ($\approx$ 2022.7) become standard references.

The **current frontier**, from roughly 2023.5 onward, is dominated by the terms that now define the agenda: *machine learning* as the umbrella term ($\approx$ 2023.4), *dropout prediction* ($\approx$ 2023.6), *academic performance* ($\approx$ 2023.5), *deep learning* ($\approx$ 2023.9) and *feature selection* ($\approx$ 2023.8), together with the most recent signals of all, *gradient boosting* ($\approx$ 2024.7), *lstm* ($\approx$ 2024.4), *predictive analytics* ($\approx$ 2024.1), *academic dropout* ($\approx$ 2024.0) and *student success* ($\approx$ 2025.0). Some classical methods such

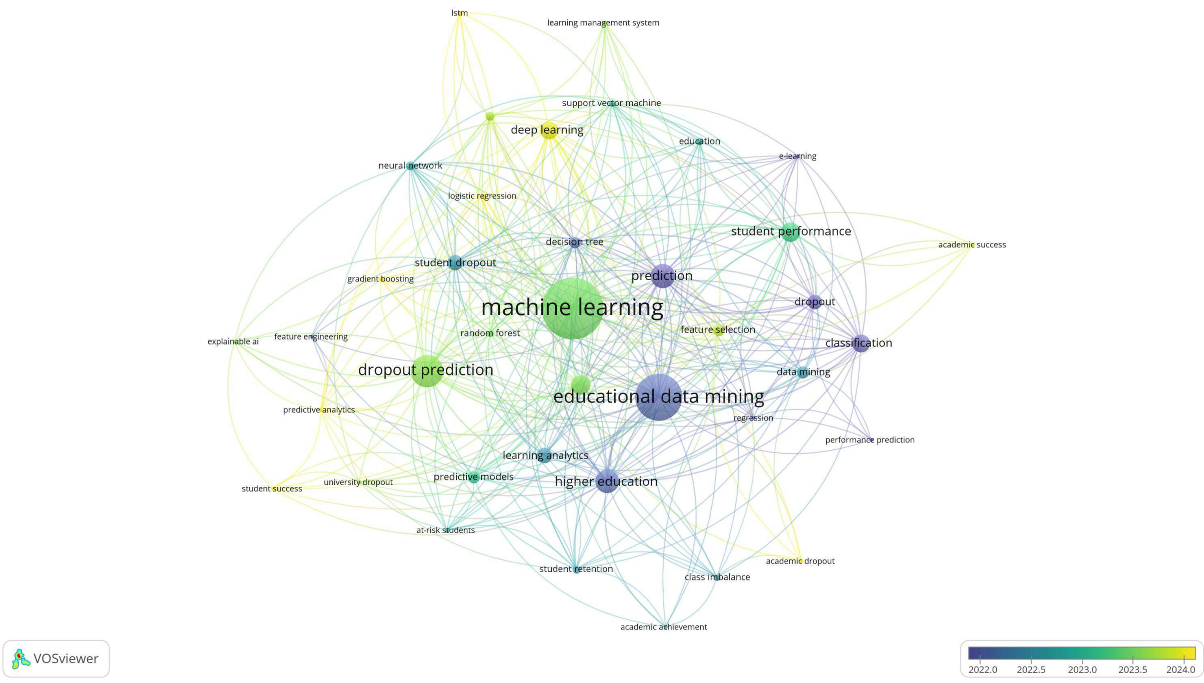

Figure 5: VOSviewer temporal overlay, with terms colored by average publication year.

as *logistic regression* (≈ 2024.0) also carry late averages: the technique is old, but it persists as a comparison baseline in recent papers, so its term age reflects continued use rather than novelty. The position of *explainable ai* (≈ 2023.6) in this leading group is notable: interpretability has moved within a few years from a secondary concern to a stated design requirement.

The overall picture is one of the field shifting from the foundational data-mining and classification vocabulary toward deep learning, advanced ensembles and explainability, while the educational substance of the problem holds steady and only the methods layered on top of it keep turning over, which is also what the most recent reviews report [12, 13, 32].

7 Discussion

The bibliometric structure described above is consistent with what the individual studies suggest, and it also makes the field's blind spots easier to see. Three points are worth drawing out, and it matters to keep separate those that rest on the map itself and those that rest on the reviews the map points to.

First, reported predictive performance is high and still rising, but it rests on fragile foundations. The headline accuracies cited in Section 2 are illustrative single-study results rather than field-level estimates, and they are not directly comparable: the 98.3% reported for one programming course was obtained with GAN- and SMOTE-based augmentation [40], and the 14% reduction in dropout was a single-institution before-and-after comparison without a control group [7]. A large share of the empirical work is built on one institution's data, which limits how far any model generalizes; the reviews we surveyed are unusually blunt on this point, observing that most studies neither test transferability across institutions nor report the conditions under which their accuracy would degrade [12, 32]. The recurrence of *class imbalance* in the term map is a reminder that even a high accuracy can mislead unless the minority, the students who actually leave, is handled with care [36, 39].

Second, the field is visibly turning toward interpretability, but the turn is incomplete. *Explainable ai* is now among the most recent terms in the network, and individual studies increasingly attach SHAP or permutation-importance analyses to otherwise opaque models [5]. Yet the reviews report that only a minority of studies treat explanation, fairness and real-world deployment as first-class requirements rather than optional extras [12, 41]. The map can support only the weaker version of this claim, namely that fairness and deployment terms are absent from or peripheral in the 39-term network; that absence motivates the concern rather than proving it. The same caution applies to any inference drawn from the map: a term's absence or marginality may reflect authors' keyword conventions, the co-occurrence threshold or the wording of the query, rather than the true state of the literature. Because a prediction here can affect a student's academic future, the gap between what these models can do technically and how responsibly they are used matters more than any of the others.

Third, very little of this work reaches the point of acting on its predictions. The reviews repeatedly note that few studies move from a risk score to a documented intervention and an evaluation of its effect [1, 2]. The exceptions are instructive: Burgos et al. coupled their model to a tutoring plan and measured a real reduction in dropout [7], and Coussement et al. designed their model around comprehensibility precisely so that retention campaigns could be personalized from it [8]. The value of prediction, these cases suggest, is realized only when it is embedded in an institutional process. The next section turns these gaps into the design of a concrete system.

8 A proposed explainable early warning system (X-EWS) framework

To be useful, the bibliometric findings have to feed back into how such systems are designed. Table 3 brings the gaps together and, for each one, gives the evidence behind it and the design direction it implies, distinguishing what is read directly off the term map from what the reviews in the corpus report. These directions point to a system whose worth is not measured by a single accuracy figure but by whether it can be explained, carried across institutions, audited for fairness, and run in practice.

Table 3: Identified gaps, the supporting evidence (from the term map or from the reviews in the corpus), and the design direction each implies.

Gap	**Evidence**	**Design direction**
Interpretability	term map: *explainable ai* appears only recently (≈ 2023.6) and stays peripheral in the network	explanations based on SHAP and LIME (Local Interpretable Model-agnostic Explanations), built into the model rather than added afterward
Transferability	reviews: single-institution datasets dominate; multi-institutional studies are rare [12, 32]	validation across cohorts and programs; federated, cross-institutional learning
Fairness	term map: fairness-related terms largely absent from the 39-term network; reviews concur [41]	fairness-aware modeling and routine subgroup auditing
Deployment	term map: operational early-warning terms are scarce; reviews: few studies evaluate interventions [2, 7]	embedding prediction in an institutional intervention-and-governance loop

Building on these directions, and on the structure that recurs across the deployed systems in the corpus [8, 10, 31], we propose a five-layer explainable early warning system, abbreviated X-EWS and shown in Figure 6. The layers are stacked so that data move upward from acquisition toward action, but governance is not a final step: it bears on every layer at once.

The layers are not invented beside the map; each descends from it. The acquisition and feature engineering layers answer Cluster 4, where the outcome variables meet feature selection and class imbalance, the data-quality problems that decide whether prediction is possible at all. The risk prediction engine draws on Cluster 2 and on the foundational vocabulary of Cluster 3, where the field keeps its classical and deep methods side by side. The explainability layer responds to one of the most recent movements inside Cluster 1, the arrival of explainable AI at the research frontier, and the intervention and governance layers answer the two gaps the map leaves most exposed, deployment and fairness.

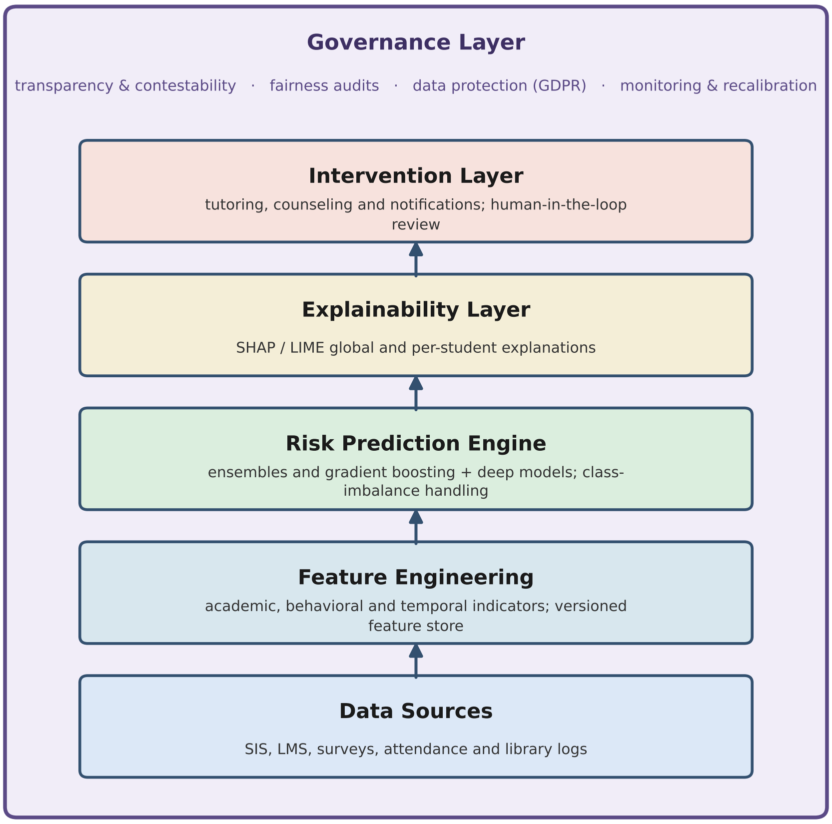

Figure 6: The proposed X-EWS framework: five functional layers, from data acquisition to intervention, wrapped by a transversal governance layer (transparency and contestability, fairness audits, data protection, monitoring and recalibration); data flow upward from acquisition toward action, while governance bears on every layer at once.

The **data acquisition layer** collects the signals that institutions already hold: enrollment and grade records from the student information system, activity logs from the learning management system, attendance, and, where available, survey and library data. Keeping acquisition explicit matters because the most transferable predictors reported in the literature, prior grades and first-year performance, come precisely from these routine sources [10, 24].

The **feature engineering layer** turns raw records into model inputs. Academic, behavioral and temporal indicators are derived and aggregated over sliding windows, and stored in a versioned feature store so that the same definitions can be reused across cohorts. Because educational data combine variables of uneven quality, principled feature selection belongs here as well, since it is often what separates a usable model from an over-fitted one [17, 23, 30].

The **risk prediction engine** is where the corpus is most developed. Ensemble and gradient-boosting models, which the literature repeatedly identifies as the strongest performers on tabular institutional data, form the backbone, complemented by deep architectures where sequential behavioral data justify them [10, 22, 39, 40]. Class imbalance is handled explicitly through resampling, since the students who leave are the minority that the model exists to find [36].

The **explainability layer** is among the elements the temporal analysis shows to be most recent and least consolidated. Rather than treating explanation as an afterthought, the framework places SHAP- or LIME-style attributions next to every risk score, both globally and for the individual student, so that an advisor can see why a student was flagged [5, 12]. The same preference for interpretable-by-design models is visible in other high-stakes screening domains, such as clinical decision support [20]. This is what allows a prediction to be questioned, justified and acted upon responsibly.

The **intervention layer** closes the loop. Risk scores are turned into action, whether that is tutoring, counseling or a timely notification, and a person always reviews the score before any decision that carries consequences for the student [7, 8]. Cutting across all five layers is a governance concern that the bibliometric gaps make non-negotiable: students must be able to see why they were flagged and to contest it; subgroups must be audited for fairness [41]; data-protection rules must be met; and the model must be monitored and recalibrated as cohorts change, so it does not silently drift. In this design the four gaps of Table 3 are addressed not in isolation but as concerns built into the architecture itself.

The framework is described at the architectural level, but each layer corresponds to a quantity that can be written down explicitly. For a student i observed up to a point t in the term, the engine produces a risk score $r(i,t) = f(x(i,t))$, where $x(i,t)$ is the feature vector assembled by the feature-engineering layer and f is the model trained in the risk-prediction layer; the student is flagged when $r(i,t)$ exceeds a program-specific threshold τ that trades recall against the volume of alerts. The explainability layer attaches to each score a vector of feature attributions $\varphi(i,t)$, for instance SHAP values, that accounts for it. Evaluation would rely on the metrics that already dominate the corpus. AUC and F1-score would measure overall discrimination, recall on the minority (dropout) class would show whether the students who leave are actually caught, and the predicted probabilities would be checked for calibration. Fairness would be assessed across protected subgroups, and a final test would ask whether the explanations in fact change what advisors do. Section 9 exercises exactly this design in a proof-of-concept on a public benchmark; full validation on institutional cohorts remains future work.

8.1 Operational data flow and deployment pipeline

The framework runs on the rhythm of the academic calendar rather than in real time. At each institutional checkpoint, at enrollment and at the end of each semester, the acquisition layer refreshes its extracts from the student information system and the learning platform, the feature layer recomputes its indicators against versioned definitions, and the risk engine scores the cohort in batch; every score above the program threshold τ becomes an alert that carries its feature attributions with it. Data, features and models are versioned together, so that any score ever produced can be reconstructed, a traceability requirement imposed by data-protection rules. Training follows the same discipline: a candidate model version is trained on past cohorts and promoted only after it clears explicit quality gates on discrimination, calibration and subgroup fairness. Deployment is staged. In shadow mode the system scores students but no one acts on it, and its outputs are compared retrospectively with observed outcomes; a canary stage then exposes alerts to a restricted group of advisors; only afterwards does the system enter production, always with a person reviewing the score before any consequential decision. Monitoring completes the cycle: input distributions are tracked with the population stability index (with the customary indicative thresholds of 0.1 and 0.25) and Kolmogorov-Smirnov tests, performance is recomputed on sliding windows, and established drift detectors [14] trigger recalibration or retraining when the population or the concept shifts, with the model registry preserving full traceability of these changes.

8.2 Fairness: evaluation metrics and mitigation

Making the fairness requirement operational demands named metrics rather than good intentions. *Demographic parity* compares alert rates across groups; *equal opportunity* compares recall on the dropout class, that is, whose actual dropouts the system misses; *equalized odds* additionally constrains false-positive rates; and *calibration within groups* asks whether a given risk score means the same observed dropout frequency in every group. For an early-warning system the equal-opportunity gap is arguably the most meaningful of these, because a missed dropout is precisely a student left without support. When an audit reveals disparities, mitigation can act before training (reweighing or resampling the data), during training (fairness-constrained objectives) or after it (group-specific decision thresholds); the last option is the most direct but also the most delicate, since differentiated treatment keyed to protected attributes raises ethical and legal questions of its own [41]. The framework therefore treats mitigation as a governed decision informed by routine audits, not as an automatic adjustment.

8.3 Implementation challenges

Several practical obstacles stand between this design and institutional operation. *Data privacy*: the system processes personal data in order to classify students by risk, so data minimization, purpose limitation, a data-protection impact assessment and a workable contestation path are preconditions, not refinements. *Institutional heterogeneity*: student information systems differ in schemas, grading scales and completeness, which is one reason models transfer poorly between institutions; federated or cross-institutional validation, as indicated in Table 3, is the corresponding design answer. *Data quality*: educational records combine variables of uneven reliability, and indicator definitions themselves change over time, which argues for the versioned feature definitions of Section 8.1. *Long-term maintenance*: models degrade silently as cohorts change, so monitoring, scheduled recalibration and governance of retraining decisions are running costs that must be budgeted from the start. *Human factors*, finally, bound everything: the alert volume implied by the threshold τ must match real advising capacity, or the system produces fatigue instead of intervention. Two further considerations are ethical rather than technical. The costs of error are asymmetric: a false negative is a student who needed support and was never flagged, while a false positive spends advising effort and, if handled carelessly, risks stigmatizing a student who was never in danger; this asymmetry is precisely why the threshold τ must remain an institutional decision rather than a technical constant. And the label itself carries weight: being marked "at risk" can shape expectations on both sides, so alerts must be framed as offers of support rather than verdicts, with the student able to see, and to contest, the reasons behind the flag.

9 Proof-of-concept instantiation on a public benchmark

To show that the architecture is more than a diagram, we instantiated its predictive core on a public benchmark: the dataset introduced by Realinho et al. [26], with 4,424 students described by 36 attributes and labeled as graduates (49.9%), dropouts (32.1%) or still enrolled (17.9%). The early-warning decision was framed as dropout versus the rest, with a stratified 70/30 train-test split, minority over-sampling applied inside the training pipeline only [36], and a gradient-boosting engine (400 trees, maximum depth 5, learning rate 0.08), consistent with the methods the corpus itself favors for tabular institutional data [10, 39].

Each layer of the framework maps onto a measurable result. The checkpoint structure of the feature layer was emulated by progressively unmasking the attributes: with only the 24 indicators known at enrollment the engine reaches an AUC of 0.82 with a dropout recall of 0.56; adding the six first-semester indicators lifts performance to 0.91 and 0.73; the full attribute set reaches 0.93 and 0.77 (Figure 7). The largest information gain arrives with the first semester, exactly where an early warning is still actionable. The intervention layer's threshold becomes concrete as well: if the institution targets a recall of at least 85%, the threshold moves to $\tau = 0.30$, catching 86% of actual dropouts at a precision of 0.78 while alerting about 35% of the cohort, a trade-off the institution can read and adjust directly. The scores are well calibrated (Brier score 0.089), so they can be treated as probabilities rather than mere rankings. The explainability layer behaves as intended: the dominant attributions, second-semester curricular units approved and tuition fees up to date, give advisors academically and financially coherent reasons for each alert, in line with the determinants reported on comparable populations [10]. The governance layer, finally, proves its necessity rather than its comfort: the fairness audit finds an equal-opportunity gap of 0.16 across age groups (recall 0.93 for students over 25, at the price of alerting 61% of them, versus 0.77 for those aged 20 or under) and lower recall for scholarship holders (0.74 versus 0.87), disparities that call for routine auditing and governed mitigation exactly as prescribed in Section 8.2.

This proof-of-concept is deliberately bounded: it uses one public benchmark, a single train-test split and the framework's predictive, explanatory and audit layers, not the full institutional loop. Its purpose is to demonstrate feasibility, namely that the predictive, explanatory and audit layers of X-EWS can all be built from off-the-shelf, auditable components; a complete empirical study, with systematic model comparisons and a full fairness audit, is being prepared as a separate paper.

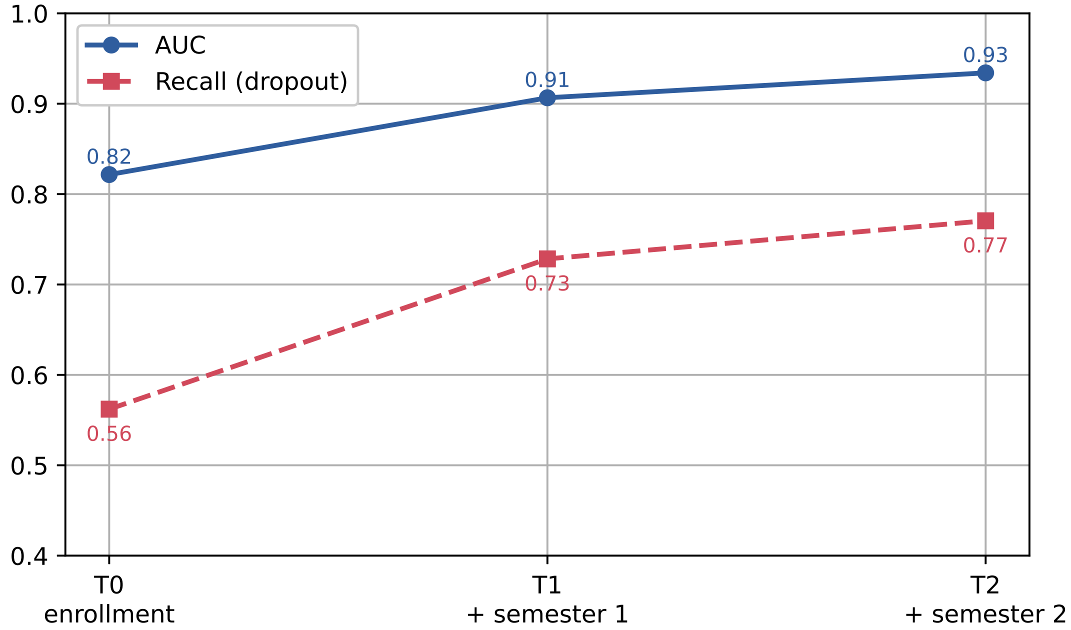

Figure 7: Early-warning checkpoints in the proof-of-concept: discrimination (AUC) and dropout recall grow as academic information accrues (T0 enrollment; T1 after semester 1; T2 after semester 2).

10 Theoretical and practical contributions

The main contribution of this paper is not a new predictive algorithm but a bibliometrically grounded system architecture that turns recurring gaps in the literature into explicit design requirements for explainable, fair and deployable early-warning systems. Theoretically, the study converts a literature previously summarized only in narrative reviews into a quantitatively mapped structure. It shows that the field organizes into four stable clusters, that its dominant vocabulary has moved from generic data mining toward deep learning and explainability, and that the same shortcomings keep returning across the reviews: interpretability stays weak, transfer between institutions is barely tested, fairness receives limited explicit attention, and relatively few predictions are deployed in practice. The X-EWS framework then ties these findings to a concrete, layered design, and the proof-of-concept in Section 9 shows that this design is not merely programmatic: its core layers can be instantiated on a public benchmark with standard components.

In practical terms, the framework offers institutions a design they can adapt, not a single algorithm to reuse unchanged. It points to the methods the evidence favors, ensembles and gradient boosting on tabular records, deep models where behavioral traces exist. Class imbalance has to be dealt with before any accuracy figure means much. And explanation, human review and data-protection governance belong at the center of the design, not at its margin. For an institutional data provider, this is what separates a one-off predictive experiment from a system that can actually be run and accounted for.

11 Conclusion

This paper examined the literature on machine learning for student dropout and academic performance prediction in higher education through a bibliometric lens, and used the result to motivate a conceptual contribution. Working from 352 Scopus-indexed documents published between 2014 and 2025, a keyword co-occurrence map produced four thematic clusters, and a temporal overlay traced the field's movement from generic data mining toward deep learning, advanced ensembles and explainable artificial intelligence.

The structure that emerges is that of a productive and methodologically mature field held back by the same recurring limitations. Most work still leans on single-institution data and shows little that it generalizes; interpretability is now named as a priority but is not yet routine practice; and the move from a risk score to a documented intervention is one that few studies actually take. Rather than leaving these as observations, we translated them into the X-EWS framework, a five-layer architecture that treats interpretability, transferability, fairness and deployment as explicit design requirements, and exercised that design in a proof-of-concept instantiation on a public benchmark. Several limitations qualify these conclusions. It rests on a single database, Scopus, on one four-term query whose wording, like any search expression, shapes the corpus it returns, and on keyword co-occurrence as its only mapping technique, so the map reported here is one reasonable reconstruction of the field, and other database, query and method choices could yield a different one; and the X-EWS framework, while exercised in a proof-of-concept on one public benchmark, still awaits validation on real institutional cohorts. Future work will address these on two fronts: extending the bibliometric analysis with citation-based techniques, such as co-citation and bibliographic coupling, and with retrieval from additional databases such as Web of Science or IEEE Xplore; and instantiating the framework on institutional cohorts, validating its predictive layers under realistic operating conditions and testing whether the explanations it produces genuinely support the interventions that follow.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

Declaration on the use of AI-assisted technologies

During the preparation of this work, the authors used AI-assisted technologies to improve the readability and language of the manuscript and to assist with formatting, figure preparation and reference management. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Alalawi, K.; Athauda, R.; Chiong, R. (2023). Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review, *Engineering Reports*, 5(12), e12699, 2023. https://doi.org/10.1002/eng2.12699

[2] Albreiki, B.; Zaki, N.; Alashwal, H. (2021). A systematic literature review of student' performance prediction using machine learning techniques, *Education Sciences*, 11(9), 552, 2021. https://doi.org/10.3390/educsci11090552

[3] Alhazmi, E.; Sheneamer, A. (2023). Early Predicting of Students Performance in Higher Education, *IEEE Access*, 11, 27579–27589, 2023. https://doi.org/10.1109/ACCESS.2023.3250702

[4] Alnasyan, B.; Basheri, M.; Alassafi, M. (2024). The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review, *Computers and Education: Artificial Intelligence*, 6, 100231, 2024. https://doi.org/10.1016/j.caeai.2024.100231

[5] Baranyi, M.; Nagy, M.; Molontay, R. (2020). Interpretable Deep Learning for University Dropout Prediction, In *SIGITE 2020 - Proceedings of the 21st Annual Conference on Information Technology Education*, 13–19, 2020. https://doi.org/10.1145/3368308.3415382

[6] Batool, S.; Rashid, J.; Nisar, M.W.; Kim, J.; Kwon, H.Y.; Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study, *Education and Information Technologies*, 28(1), 905–971, 2023. https://doi.org/10.1007/s10639-022-11152-y

[7] Burgos, C.; Campanario, M.L.; Peña, D.D.L.; Lara, J.A.; Lizcano, D.; Martínez, M.A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout, *Computers and Electrical Engineering*, 66, 541–556, 2018. https://doi.org/10.1016/j.compeleceng.2017.03.005

[8] Coussement, K.; Phan, M.; De Caigny, A.; Benoit, D.F.; Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model, *Decision Support Systems*, 135, 113325, 2020. https://doi.org/10.1016/j.dss.2020.113325

[9] De-La-Cruz, P.; Rojas-Coaquira, R.; Vega-Huerta, H.; Pérez-Quintanilla, J.; Lagos-Barzola, M. (2022). A Systematic Review Regarding the Prediction of Academic Performance, *Journal of Computer Science*, 18(12), 1219–1231, 2022. https://doi.org/10.3844/JCSSP.2022.1219.1231

[10] Delogu, M.; Lagravinese, R.; Paolini, D.; Resce, G. (2024). Predicting dropout from higher education: Evidence from Italy, *Economic Modelling*, 130, 106583, 2024. https://doi.org/10.1016/j.econmod.2023.106583

[11] Devasia, T.; Vinushree, T.P.; Hegde, V. (2016). Prediction of students performance using Educational Data Mining, In *Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016*, 7684167, 2016. https://doi.org/10.1109/SAPIENCE.2016.7684167

[12] Duro, B.; Gomes, A.; Correia, F.B.; Borges, A.R.; Bernardino, J. (2026). Machine Learning and Deep Learning for Dropout Prediction in Higher Education: A Review, *Computers*, 15(3), 164, 2026. https://doi.org/10.3390/computers15030164

[13] Fierro Saltos, W.R.; Fierro Saltos, F.E.; Elizabeth Alexandra, V.S.; Rivera Guzmán, E.F. (2025). Leveraging Artificial Intelligence for Sustainable Tutoring and Dropout Prevention in Higher Education: A Scoping Review on Digital Transformation, *Information (Switzerland)*, 16(9), 819, 2025. https://doi.org/10.3390/info16090819

[14] Gama, J.; Žliobaitė, I.; Bifet, A.; Pechenizkiy, M.; Bouchachia, A. (2014). A survey on concept drift adaptation, *ACM Computing Surveys*, 46(4), 44, 2014. https://doi.org/10.1145/2523813

[15] Guanin-Fajardo, J.H.; Guaña-Moya, J.; Casillas, J. (2024). Predicting Academic Success of College Students Using Machine Learning Techniques, *Data*, 9(4), 60, 2024. https://doi.org/10.3390/data9040060

[16] He, Y.; Chen, R.; Li, X.; Hao, C.; Liu, S.; Zhang, G.; et al. (2020). Online at-risk student identification using RNN-GRU joint neural networks, *Information (Switzerland)*, 11(10), 474, 2020. https://doi.org/10.3390/info11100474

[17] Kabathova, J.; Drlik, M. (2021). Towards predicting student's dropout in university courses using different machine learning techniques, *Applied Sciences (Switzerland)*, 11(7), 3130, 2021. https://doi.org/10.3390/app11073130

[18] Kocsis, Á.; Molnár, G. (2025). Factors influencing academic performance and dropout rates in higher education, *Oxford Review of Education*, 51(3), 414–432, 2025. https://doi.org/10.1080/03054985.2024.2316616

[19] Lopez Guarin, C.E.; Guzman, E.L.; Gonzalez, F.A. (2015). A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining, *Revista Iberoamericana de Tecnologias del Aprendizaje*, 10(3), 7156098, 2015. https://doi.org/10.1109/RITA.2015.2452632

[20] Lv, Y.; Weng, J.; Li, J.; Chen, W.; Huang, H.; Zhao, Y. (2025). A New Evaluation Model for Traumatic Severe Pneumothorax Based on Interpretable Machine Learning, *International Journal of Computers Communications & Control*, 20(1), 6830, 2025. https://doi.org/10.15837/ijccc.2025.1.6830

[21] Monteverde-Suárez, D.; González-Flores, P.; Santos-Solórzano, R.; García-Minjares, M.; Zavala-Sierra, I.; de la Luz, V.L.; et al. (2024). Predicting students' academic progress and related attributes in first-year medical students: an analysis with artificial neural networks and Naïve Bayes, *BMC Medical Education*, 24(1), 74, 2024. https://doi.org/10.1186/s12909-023-04918-6

[22] Nabil, A.; Seyam, M.; Abou-Elfetouh, A. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks, *IEEE Access*, 9, 140731–140746, 2021. https://doi.org/10.1109/ACCESS.2021.3119596

[23] Nachouki, M.; Mohamed, E.A.; Mehdi, R.; Abou Naaj, M. (2023). Student course grade prediction using the random forest algorithm: Analysis of predictors' importance, *Trends in Neuroscience and Education*, 33, 100214, 2023. https://doi.org/10.1016/j.tine.2023.100214

[24] Ortiz-Lozano, J.M.; Rua-Vieites, A.; Bilbao-Calabuig, P.; Casadesús-Fa, M. (2020). University student retention: Best time and data to identify undergraduate students at risk of dropout, *Innovations in Education and Teaching International*, 57(1), 74–85, 2020. https://doi.org/10.1080/14703297.2018.1502090

[25] Rastrollo-Guerrero, J.L.; Gómez-Pulido, J.A.; Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review, *Applied Sciences (Switzerland)*, 10(3), 1042, 2020. https://doi.org/10.3390/app10031042

[26] Realinho, V.; Machado, J.; Baptista, L.; Martins, M.V. (2022). Predicting Student Dropout and Academic Success, *Data*, 7(11), 146, 2022. https://doi.org/10.3390/data7110146

[27] Rizwan, S.; Nee, C.K.; Garfan, S. (2025). Identifying the Factors Affecting Student Academic Performance and Engagement Prediction in MOOC Using Deep Learning: A Systematic Literature Review, *IEEE Access*, 13, 18952–18982, 2025. https://doi.org/10.1109/ACCESS.2025.3533915

[28] Rodríguez-Ortiz, M.Á.; Santana-Mancilla, P.C.; Anido-Rifón, L.E. (2025). Machine Learning and Generative AI in Learning Analytics for Higher Education: A Systematic Review of Models, Trends, and Challenges, *Applied Sciences (Switzerland)*, 15(15), 8679, 2025. https://doi.org/10.3390/app15158679

[29] Rovira, S.; Puertas, E.; Igual, L. (2017). Data-driven system to predict academic grades and dropout, *PLoS ONE*, 12(2), e0171207, 2017. https://doi.org/10.1371/journal.pone.0171207

[30] Roy, K.; Farid, D.M. (2024). An Adaptive Feature Selection Algorithm for Student Performance Prediction, *IEEE Access*, 12, 75577–75598, 2024. https://doi.org/10.1109/ACCESS.2024.3406252

[31] Salloum, S.A.; Basiouni, A.; Alfaisal, R.; Salloum, A.; Shaalan, K. (2024). Predicting Student Retention in Higher Education Using Machine Learning, In *Communications in Computer and Information Science*, 2162 CCIS, 197–206, 2024. https://doi.org/10.1007/978-3-031-65996-6_17

[32] Shi, H.; Zhang, N.; Caskurlu, S.; Na, H. (2025). Applications of Machine Learning for at-Risk Student Prediction in Online Education: A 10-Year Systematic Review of Literature, *Journal of Computer Assisted Learning*, 41(4), e70058, 2025. https://doi.org/10.1111/jcal.70058

[33] Sinval, J.; Oliveira, P.; Novais, F.; Almeida, C.M.; Telles-Correia, D. (2025). Exploring the impact of depression, anxiety, stress, academic engagement, and dropout intention on medical students' academic performance: A prospective study, *Journal of Affective Disorders*, 368, 665–673, 2025. https://doi.org/10.1016/j.jad.2024.09.116

[34] Stasolla, F.; Zullo, A.; Maniglio, R.; Passaro, A.; Di Gioia, M.; Curcio, E.; et al. (2025). Deep Learning and Reinforcement Learning for Assessing and Enhancing Academic Performance in University Students: A Scoping Review, *AI (Switzerland)*, 6(2), 40, 2025. https://doi.org/10.3390/ai6020040

[35] Tan, M.; Shao, P. (2015). Prediction of student dropout in E-learning program through the use of machine learning method, *International Journal of Emerging Technologies in Learning*, 10(1), 11–17, 2015. https://doi.org/10.3991/ijet.v10i1.4189

[36] Thammasiri, D.; Delen, D.; Meesad, P.; Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition, *Expert Systems with Applications*, 41(2), 321–330, 2014. https://doi.org/10.1016/j.eswa.2013.07.046

[37] Umer, R.; Susnjak, T.; Mathrani, A.; Suriadi, S. (2017). On predicting academic performance with process mining in learning analytics, *Journal of Research in Innovative Teaching and Learning*, 10(2), 160–176, 2017. https://doi.org/10.1108/JRIT-09-2017-0022

[38] Vacarescu, S.P.I.; Balas, V.E.; Paraschiv, N.; Petcut, F. (2025). A Systematic Review for Quality within the Automotive Industry 4.0, *International Journal of Computers Communications & Control*, 20(4), 7123, 2025. https://doi.org/10.15837/ijccc.2025.4.7123

[39] Villar, A.; de Andrade, C.R.V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study, *Discover Artificial Intelligence*, 4(1), 2, 2024. https://doi.org/10.1007/s44163-023-00079-z

[40] Vives, L.; Cabezas, I.; Vives, J.C.; Reyes, N.G.; Aquino, J.; Condor, J.B.; et al. (2024). Prediction of Students' Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks, *IEEE Access*, 12, 5882–5898, 2024. https://doi.org/10.1109/ACCESS.2024.3350169

[41] Yu, R.; Lee, H.; Kizilcec, R.F. (2021). Should College Dropout Prediction Models Include Protected Attributes?, In *L@S 2021 - Proceedings of the 8th ACM Conference on Learning @ Scale*, 91–100, 2021. https://doi.org/10.1145/3430895.3460139

Cite this paper as:

Radulescu, D.A.; Balas, V.E. (2026). Explainable Early Warning Systems for Student Dropout Prediction: Insights from a Bibliometric Analysis, *International Journal of Computers Communications & Control*, 21(4), 7599, 2026.
https://doi.org/10.15837/ijccc.2026.4.7599